JOURNAL OF NUMBER THEORY **66**, 211–224 (1997)
ARTICLE NO. NT972169

On Infinite Sum-free Sets of Natural Numbers

Tomasz Łuczak and Tomasz Schoen*

Department of Discrete Mathematics, Adam Mickiewicz University, Poznań, Poland

Communicated by Alan C. Woods

Received December 12, 1995

A subset of the natural numbers is k-sum-free if it contains no solutions of the equation $x_1 + \cdots + x_k = y$, and strongly k-sum-free when it is ℓ-sum-free for every $\ell = 2, ..., k$. It is shown that every k-sum-free set with upper density larger than $1/(k+1)$ is a subset of a periodic k-sum-free set and that each k-sum-free set with

View metadata, citation and similar papers at core.ac.uk

studied also for strongly k-sum-free sets. © 1997 Academic Press

1. INTRODUCTION

A subset A of a semigroup is *k-sum-free* for $k \geqslant 2$ if $x_1 + x_2 + \cdots + x_k \notin A$ for all $x_1, x_2, ..., x_k \in A$, and it is *strongly k-sum-free* if it is ℓ-sum-free for every $\ell = 2, ..., k$. In this note we study the structure of large infinite k-sum-free subsets of the natural numbers where as the measure of the size of a set $A \subseteq \mathbb{N}$ we use its upper density, defined as

$$\bar{\mathrm{d}}(A) = \limsup_{n \to \infty} \bar{\mathrm{d}}_n(A), \qquad \text{where} \quad \bar{\mathrm{d}}_n(A) = \frac{|A \cap \{1, ..., n\}|}{n}.$$

One can immediately notice that the upper density of a strongly k-sum-free set does not exceed $1/k$. On the other hand, the set of odd numbers is k-sum-free for arbitrary large even k. Calkin and Erdős [1] conjectured that the maximum density of k-sum-free sets depends mainly on parity conditions of such type and for every k-sum-free set A we have $\bar{\mathrm{d}}(A) \leqslant 1/\rho_1(k)$, where

$$\rho_1(k) = \min\{i : i \nmid k-1\}.$$

We settle their conjecture in the affirmative.

* The authors thank U.S.–Polish Research Grant 93113003 and KBN Grant 2 P03A 023 09 for their generous support.

0022-314X/97 $25.00
Copyright © 1997 by Academic Press
All rights of reproduction in any form reserved.

Clearly, both bounds $1/k$ and $1/\rho_1(k)$ are sharp, but one can ask whether it is possible to specify all "extremal" k-sum-free and strongly k-sum-free sets which achieve this density. In the case of sum-free sets (i.e., when $k=2$) it is known that the set of odd numbers is the only maximal sum-free set in a rather strong sense: each sum-free set with density larger than 2/5 consists of odd numbers only and the example of the sum-free set $\{n: n \equiv 2, 3 \pmod 5\}$ shows that 2/5 cannot be replaced by a smaller constant (see Łuczak [4] or Deshouillers, Freiman, Sós, and Tamkin [2], where a much stronger finite version of this fact is proved). We study the analogous problem for $k \geqslant 3$ showing that each strongly k-sum-free set with upper density larger than $1/(k+1)$ is a subset of a set $\{n: n \equiv s \pmod k\}$ for some s such that $(s,k)=1$. For k-sum-free sets a somewhat stronger result is shown: we prove that every k-sum-free set with upper density larger than $i/[i(k+1)-k+1]$ for some $i \geqslant 1$ is a subset of a k-sum-free set which is a union of at most $i-1$ arithmetic progressions with the same difference. Thus, in particular, for $k \geqslant 6$, each k-sum-free set with upper density larger than $1/\rho_2(k)$, where

$$\rho_2(k) = \min\{i > \rho_1(k) : i \nmid k-1\},$$

is a subset of an arithmetic progression with difference $\rho_1(k)$.

Since all [strongly] k-sum-free of large density mentioned so far are subsets of [strongly] k-sum-free sets which are unions of a finite number of arithmetic progressions it seems natural to ask about the minimum density which forces a [strongly] k-sum-free set to be of such a regular structure. The complete answer for this question is provided by Theorems 1 and 1* where it is shown that every [strongly] k-sum-free set with upper density larger than $1/(k+1)$ [$1/(2k-1)$] is a subset of a [strongly] k-sum-free set which is a union of a finite number of arithmetic progressions of the same difference each. Moreover, neither $1/(k+1)$ nor $1/(2k-1)$ can be replaced by a smaller constant. As a matter of fact these two results are crucial for our argument since they show that properties of dense [strongly] k-sum-free subsets of the natural numbers can be deduced from properties of [strongly] k-sum-free subsets of finite cyclic groups.

2. k-SUM-FREE SETS

We start with a result which, roughly speaking, states that dense k-sum-free subsets of the natural numbers containing long arithmetic progressions of difference m do not contain solutions of some special class of equations $x_1 + \cdots + x_k = y \pmod m$.

For a set $A \subseteq \mathbb{N}$ define

$$\mathrm{Df}_k(A) = \{u - v_1 - \cdots - v_{k-1} : u, v_1, ..., v_{k-1} \in A\}$$

and

$$\mathrm{Df}_{\leqslant k}(A) = \bigcup_{\ell=2}^{k} \mathrm{Df}_\ell(A).$$

Thus, for a k-sum-free set A we have $A \cap \mathrm{Df}_k(A) = \varnothing$ and each strongly k-sum-free set A fulfills $A \cap \mathrm{Df}_{\leqslant k}(A) = \varnothing$.

The following Lemma 1, as well as Lemma 1* stated in Section 3, can be viewed as a generalization of a similar result from [4], where the case $k = 2$ and $m = 1$ was considered.

LEMMA 1. *Let $A \subseteq \mathbb{N}$ be a k-sum-free set and $x, x+m, ..., x+(i-1)\,m \in A$. If there exists $d \in \mathrm{Df}_k(A)$ such that $d \equiv x \pmod{m}$ then*

$$\overline{\mathrm{d}}(A) \leqslant \frac{k+i-2}{(k+1)\,i+k-3}.$$

Proof. Let $A \subseteq \mathbb{N}$ be a k-sum-free set and let $x, x+m, ..., x+(i-1)\,m \in A$, where $d \equiv x \pmod{m}$ for some $d \in \mathrm{Df}_k(A)$.

We consider first the case when $d < x$. Because A is k-sum-free we have $x, x+m, ..., x+(i-1)\,m \notin \mathrm{Df}_k(A)$ and so there exists $d_1 = u - v_1 - \cdots - v_{k-1} \in \mathrm{Df}_k(A)$ such that $u, v_1, ..., v_{k-1} \in A$ and $d_1 + m, d_1 + 2m, ..., d_1 + im \notin \mathrm{Df}_k(A)$. Let us put

$$C = \{a \in A : a + jm \notin A \text{ for every } f = 1, 2, ..., m\}.$$

Then, sets

$$A,\ A-(k-1)\,x,\ C+m, ..., C+(i-1)\,m$$

are disjoint, and thus, because all of them are shifted copies of either A or $C \subseteq A$, for every n large enough we get

$$2\,\overline{\mathrm{d}}_n(A) + (i-1)\,\overline{\mathrm{d}}_n(C) \leqslant 1 + O(1/n),$$

and

$$\overline{\mathrm{d}}_n(C) \leqslant (1 - 2\,\overline{\mathrm{d}}_n(A) + O(1/n))/(i-1). \qquad (*)$$

On the other hand, consider the family of $k+1$ sets A, $(A \backslash C) + v_1 + \cdots + v_{k-1}$, $(A \backslash C) + u + v_1 + \cdots + v_{k-2}, ...,$ $(A \backslash C) + (k-2)\,u + v_1$, $A + (k-1)u$. It is not hard to see that all these sets are pairwise disjoint. Indeed, since A is k-sum-free it shares no elements with any other set from the family. Furthermore, suppose that for some s, t, such that $0 \leqslant s < t \leqslant k-1$, we have

$$((A \backslash C) + su + v_1 + \cdots + v_{k-s-1}) \cap (A + tu + v_1 + \cdots + v_{k-t-1}) \neq \varnothing.$$

Then one can find $a \in A \backslash C$ and $b \in A$ such that

$$a + su + v_1 + \cdots + v_{k-s-1} = b + tu + v_1 + \cdots + v_{k-t-1},$$

so

$$\begin{aligned} a - b - (t-s-1)\, u - v_1 - \cdots - v_{k-t-1} - v_{k-s} - \cdots - v_{k-1} \\ = u - v_1 - \cdots - v_{k-1}, \end{aligned}$$

and, consequently

$$d_1 = a - b - (t-s-1)\, u - v_1 - \cdots - v_{k-t-1} - v_{k-s} - \cdots - v_{k-1} \in \mathrm{Df}_k(A).$$

Furthermore, since $a \in A \backslash C$, there exists j_0 such that $1 \leqslant j_0 \leqslant i$ and $a + j_0 m \in A$. But then also $d_1 + j_0 m \in \mathrm{Df}_k(A)$, contradicting the choice of d_1.

Thus, we arrive at

$$2\, \overline{\mathrm{d}}_n(A) + (k-1)(\overline{\mathrm{d}}_n(A) - \overline{\mathrm{d}}_n(C)) \leqslant 2\, \overline{\mathrm{d}}_n(A) + (k-1)\, \overline{\mathrm{d}}_n(A \backslash C) \leqslant 1 + O(1/n),$$

which, together with $(*)$, gives

$$\overline{\mathrm{d}}_n(A) \leqslant \frac{k+i-2}{(k+1)\, i + k - 3} + O(1/n),$$

and so leads to the required upper bound for $\overline{\mathrm{d}}(A)$.

The case when $d > x$ can be handle in a very similar manner only now one must replace d_1 by $d_2 \in \mathrm{Df}_k(A)$ such that $d_2 - m, d_2 - 2m, ..., d_2 - im \notin \mathrm{Df}_k(A)$, and instead of C consider the set

$$C' = \{a \in A : a - jm \notin A \text{ for every } j = 1, 2, ..., m\}. \quad \blacksquare$$

The crucial information on the structure of dense k-sum-free sets is given by Theorem 1 below, which says that every k-sum-free set with upper density larger than $1/(k+1)$ is contained in a "periodic" k-sum-free set, and thus naturally corresponds to a k-sum-free subset of finite cyclic group. We deduce this fact from Lemma 1 and the following consequence of a finite version of Szemerédi's density theorem [8].

LEMMA 2. *Let $\varepsilon > 0$ and $i \in N$ be arbitrary. Then there exists a constant $L(\varepsilon, i)$ such that for every set $A \subseteq N$ with $\bar{\mathrm{d}}(A) > \varepsilon$ we have $x, x+m, ..., x+(i-1)\, m \in A$ for some $x, m \in N$ and $m \leqslant L(\varepsilon, i)$.*

THEOREM 1. *For every $k \geqslant 2$ and $\varepsilon > 0$ there exists a natural number $M = M(k, \varepsilon)$ such that each k-sum-free set of upper density larger than $1/(k+1) + \varepsilon$ is contained in a k-sum-free set which is a union of arithmetic progressions of difference M each.*

Proof. Let $\varepsilon > 0$ and let $A \subseteq \mathbb{N}$ be a k-sum-free set with density $\bar{\mathrm{d}}(A) > 1/(k+1) + \varepsilon$. Choose $i_0 \in \mathbb{N}$ in such a way that

$$\bar{\mathrm{d}}(A) > \frac{k + i_0 - 2}{(k+1)\, i_0 + k - 3},$$

and set $M = L!$, where $L = L(\varepsilon, i_0)$ is the constant specified in Lemma 2. Moreover, set

$$R = \{ r \in N \colon \text{there exists } a \in A, \text{ such that } r \equiv a \pmod{M} \}$$

and

$$D = \{ r \in R \colon r + t_1 + \cdots + t_{k-1} \notin R \text{ for each } t_1, ..., t_{k-1} \in R \}.$$

Now, suppose that A is contained in no k-sum-free set which is a union of a finite number of arithmetic progression of difference M. Then R is not k-sum-free and for some $x, x_1, ..., x_k \in R$ we have $x_1 + \cdots + x_k = x$. Let us consider $k+1$ sets D, $D + x_1 + \cdots + x_{k-1}$, $D + x + x_1 + \cdots + x_{k-2}$, ..., $D + (k-1)\, x$. Note that the fact that D is k-sum-free implies that all above sets are disjoint. Indeed, suppose that for some i, j, where $0 \leqslant i < j \leqslant k-1$, we have

$$(D + ix + x_1 + \cdots + x_{k-i-1}) \cap (D + jx + x_1 + \cdots + x_{k-j-1}) \neq \varnothing.$$

Then there exist $a, b \in D$ such that

$$a + ix + x_1 + \cdots + x_{k-i-1} = b + jx + x_1 + \cdots + x_{k-j-1},$$

so that

$$\begin{aligned} a &= (j-i)\, x + b - x_{k-j} - \cdots - x_{k-i-1} \\ &= (j-i-1)\, x + b + x_1 + \cdots + x_{k-j-1} + x_{k-i} + \cdots + x_k. \end{aligned}$$

The above equality contradicts the fact that D is k-sum-free, so $\bar{\mathrm{d}}(D) \leqslant 1/(k+1)$ and $\bar{\mathrm{d}}(A \backslash D) > \varepsilon$.

From Szemerédi's theorem it follows that there are numbers $m \leqslant L$ and $x \in N$ such that $x, x+m, ..., x+(i_0-1)\, m \in A \backslash D$, so $x, x+m, ..., x+(i_0-1)\, m \in R \backslash D$. Moreover, using the definition of sets D and R, one can find $t, t_1, ..., t_{k-1} \in R$ such that $x+t_1+\cdots+t_{k-1}=t$, and some $u, v_1, ..., v_{k-1} \in A$ satisfy relations $u \equiv t \pmod M$ and $v_i \equiv t_i \pmod M$ for every $i=1, 2, ..., k-1$.

Now set $d=u-v_1-\cdots-v_{k-1}$, so that $d \in \mathrm{Df}_k(A)$ and $d \equiv x \pmod M$. Since M is divisible by m we have $d \equiv x \pmod m$ and $x, x+m, ..., x+(i_0-1)\, m \in A$. Thus, Lemma 1 gives

$$\bar{\mathrm{d}}(A)<\frac{k+i_0-2}{(k+1)\, i_0+k-3}$$

contradicting the choice of A. ∎

Corollary. *For every $k \geqslant 2$ and $\varepsilon>0$ there exists $M=M(k, \varepsilon)$ such that for every k-sum-free set $A \subseteq \mathbb{N}$ which contains multiplicity of n for every $n=1, 2, ..., M$, we have $\bar{\mathrm{d}}(A) \leqslant 1/(k+1)+\varepsilon$. In particular, if a k-sum-free set $A \subseteq \mathbb{N}$ contains multiplicities of every natural number then $\bar{\mathrm{d}}(A) \leqslant 1/(k+1)$.*

Let us remark that the constant $1/(k+1)$ in Theorem 1 and the corollary cannot be replaced by a smaller one. Indeed, following Erdős [3] for an irrational number α we set

$$E_{\alpha, k}=\left\{n \in N: \alpha n-[\alpha n] \in\left(\frac{1}{k^2-1}, \frac{k}{k^2-1}\right)\right\}.$$

Then $E_{\alpha, k}$ is a k-sum-free set of density $1/(k+1)$, and it is not hard to check that it contains multiplicities of every natural number and thus is not contained in any k-sum-free set which is a finite union of arithmetic progressions.

Theorem 1 states that each k-sum-free set $A \subseteq \mathbb{N}$ with $\bar{\mathrm{d}}(A) \geqslant 1/(k+1)$ is contained in some set $B=\{n: n \equiv a_1, ..., a_j \pmod M\}$ for some k-sum-free subset $B'=\{a_1, ..., a_j\}$ of a cyclic group $\mathbb{Z}_M$. We shall use this connection and apply Kneser's well-known lower bound for the size of the sum of sets in abelian groups to obtain more precise characterization of the structure of dense k-sum-free sets. Let us recall that the *stabilizer* of subset A of a group G, denoted by $\Gamma(A)$, is the maximal subgroup H of G which leaves A invariant, i.e. for which $A+H=A$. Then Kneser's theorem can be stated as follows.

Lemma 3. *If $A_1, ..., A_k$ are subsets of a finite abelian group G then*

$$|A_1+\cdots+A_k| \geqslant |A_1|+\cdots+|A_k|-(k-1)\, \Gamma(A_1+\cdots+A_k).$$

In our argument we shall also need the following elementary fact on the stabilizers of maximal k-sum-free sets. For simplicity of notation for a group G, a natural number ℓ and $a \in G$, $A \subseteq G$ we write

$$\ell a = \underbrace{a + \cdots + a}_{\ell \text{ times}}, \qquad \ell A = \underbrace{A + \cdots + A}_{\ell \text{ times}}.$$

Claim 1. *If A is a maximal k-sum-free subset of a finite abelian group G then $\Gamma(kA) = \Gamma(A)$.*

Proof. Clearly, $\Gamma(A) \subseteq \Gamma(kA)$, so it is enough to check that $\Gamma(kA) \subseteq \Gamma(A)$. Suppose that this is not the case and let $g \in \Gamma(kA) \setminus \Gamma(A)$ and $a \in A$ be such that $a + g \notin A$. Note that then for every ℓ, $1 \leqslant \ell \leqslant k$, and $a_1, ..., a_k \in A$ we have

$$\ell(a+g) + a_{\ell+1} + \cdots + a_k = \ell g + (\ell a + a_{\ell+1} + \cdots + a_k) \in kA$$

and thus $\ell(a+g) + a_{\ell+1} + \cdots + a_k \notin A$. Furthermore, either of relations $\ell(a+g) + a_{\ell+1} + \cdots + a_k = a + g$ and $a_1 + \cdots + a_k = a + g$ implies that $a \in kA$ which contradicts the fact that a belongs to the k-sum-free set A. Thus, the set $A \cup \{a+g\}$ is k-sum-free contradicting the maximality of A. ∎

Now Theorem 1 can be strengthened in the following way.

Theorem 2. *Let $A \subseteq \mathbb{N}$ be a k-sum-free set such that*

$$\bar{\mathrm{d}}(A) > \frac{i}{i(k+1) - k + 1}$$

for some natural $i \geqslant 2$. Then A is contained in a k-sum-free set which is a union of at most $i-1$ arithmetic progressions of the same difference. In particular, each k-sum-free set $A \subseteq \mathbb{N}$ with $\bar{\mathrm{d}}(A) > 2/(k+3)$ is contained in a k-sum-free arithmetic progression.

Proof. Let $A \subseteq \mathbb{N}$ be a k-sum-free set with upper density larger than $i/[i(k+1) - k + 1]$ for some $i \geqslant 2$. Since $i/[i(k+1) - k + 1] > 1/(k+1)$, from Theorem 1 it follows that there exists M and $1 \leqslant a_1 < \cdots < a_j < M$ such that A is contained in $B = \{n : n \equiv a_1, ..., a_j \pmod{M}\}$ and $B' = \{a_1, ..., a_j\}$, viewed as a subset of cyclic group $\mathbb{Z}_M$, is k-sum-free. Let us choose the smallest M with this property. Then the stabilizer of B' in $\mathbb{Z}_M$ is trivial, and, due to Claim 1, so is the stabilizer of kB'. Thus Kneser's results gives us

$$|kB'| \geqslant k\,|B'| - (k-1)$$

and since $(kB') \cap B' = \varnothing$,

$$M \geqslant |B'| + |kB'| \geqslant (k+1)\,|B'| - (k-1) = j(k+1) - k + 1.$$

Thus

$$\frac{i}{i(k+1)-k+1} < \overline{\mathrm{d}}(A) \leqslant \overline{\mathrm{d}}(B) = \frac{j}{M} \leqslant \frac{j}{j(k+1)-k+1},$$

and consequently $j < i$. ▮

For many pairs of i and k the result above is sharp, although, in general case, this fact depends heavily on the arithmetic properties of i and k. For instance, one can check that if $i = 2$, $k \not\equiv 1 \pmod 4$ and ℓ is a solution of the equation $4\ell + 2 \equiv 0 \pmod{(k+3)}$, then the set $A_{2,k} = \{n : n \equiv \ell, \ell + 1 \pmod{(k+3)}\}$ is k-sum-free and $\overline{\mathrm{d}}(A_{2,k}) = 2/(k+3)$. Let us also remark that in a special case when $k = 2$ a stronger result can be deduced from Kemperman's result [5] (which, in a way, supplements Kneser's theorem): it turns out that every maximal sum-free subset of the natural numbers of upper density larger than 1/3 is of the form $\{n : n \equiv ji, j(i+1), ..., j(2i-1) \pmod{(3i-1)}\}$ for some natural i and j (see [7] for more details).

Theorem 2 provides the immediate answer for the question of Calkin and Erdős mentioned in the Introduction.

Corollary 1. *For every k-sum-free set $A \subseteq \mathbb{N}$ we have $\overline{\mathrm{d}}(A) \leqslant 1/\rho_1(k)$, where*

$$\rho_1(k) = \min\{i : i \nmid k - 1\}.$$

Proof. Let $A \subseteq \mathbb{N}$ be a maximal k-sum-free set with $\overline{\mathrm{d}}(A) > 1/\rho_1(k)$. Since $1/\rho_1(k) \geqslant 2/(k+3)$ Theorem 2 implies that $A = \{n : n \equiv q \pmod M\}$ for some $M < \rho_1(k)$ and a. Nonetheless, from the definition of $\rho_1(k)$ it follows that M divides $k-1$ and so $ka \equiv a \pmod M$ which contradicts the fact that A is k-sum-free. ▮

We have already mentioned that the set of odd natural numbers is the only maximal sum-free set with upper density larger than 2/5. In the following result we list all such "extremal" k-sum-free sets for $k \geqslant 3$.

Corollary 2. *Let A be a k-sum-free set of natural numbers.*

(i) *If $k = 3$ and $\overline{\mathrm{d}}(A) > 2/7$ then A is a subset of one of four following sets: $\{n : n \equiv 1 \pmod 3\}$, $\{n : n \equiv 1 \pmod 3\}$, $\{n : n \equiv 1, 2 \pmod 6\}$ and $\{n : n \equiv 5, 6 \pmod 6\}$. Furthermore, 3-sum-free set $\{n : n \equiv 1, 2 \pmod 7\}$ shows that 2/7 cannot be replaced by any smaller constant.*

(ii) *If* $k=4$ *and* $\overline{\mathrm{d}}(A)>2/7$ *then* $A\subseteq\{n: n\equiv 1 \pmod 2\}$, *and* $2/7$ *cannot be replaced by any smaller constant because of the* 4*-sum-free set* $\{n: n\equiv 3, 4 \pmod 7\}$.

(iii) *If* $k=5$ *and* $\overline{\mathrm{d}}(A)>2/9$, *then* $A\subseteq\{n: n\equiv s \pmod 3\}$, *for some* $1\leqslant s\leqslant 2$. *Furthermore, the* 5*-sum-free set* $\{n: n\equiv 3, 4 \pmod 9\}$ *shows that this estimate is sharp.*

(iv) *If* $k\geqslant 6$ *and* $\overline{\mathrm{d}}(A)>1/\rho_2(k)$, *where*

$$\rho_2(k)=\min\{i>\rho_1(k): i\nmid k-1\},$$

then $A\subseteq\{n: n\equiv s \pmod{\rho_1(k)}\}$, *where the numbers* s *and* $\rho_1(k)$ *are relatively prime.*

Proof. Let $k=3$. From Theorem 2 it follows that if for 3-sum-free set A we have $\overline{\mathrm{d}}(A)>4/(4\cdot 4-2)=2/7$ then A is contained in a 3-sum-free set B which is a union of at most three arithmetic progressions of the same difference, say M. Thus, since $\overline{\mathrm{d}}(B)>2/7$, we have $M\leqslant 10$ and the fact that, due to Corollary 1, $\overline{\mathrm{d}}(B)\leqslant 1/3$ reduces the number of cases we need to examine even further. An easy direct check shows that there are two 3-sum-free arithmetic progressions of difference 3, $\{n: n\equiv 1, 2 \pmod 6\}$ and $\{n: n\equiv 5, 6 \pmod 6\}$ are only 3-sum-free sets which are unions of two arithmetic progressions of difference 6 and there are no 3-sum-free sets which are unions of three arithmetic progressions of difference 9 or 10 each. This completes the proof of (i). The case $k=4$ follows immediately from Theorem 2 applied for $i=2$ and the fact that no arithmetic progression of difference 3 is 4-sum-free. Similarly, Theorem 2 implies that a 5-sum-free sets A with density larger than $2/9>3/(3\cdot 6-4)$ is contained in a 5-sum-free union of at most two arithmetic progression, and no arithmetic progression of difference 4 is 5-sum-free and no 5-sum-free set is a sum of two arithmetic progressions of difference 8. Finally, in order to complete the proof let us notice that if $k\geqslant 6$ then $1/\rho_2(k)\geqslant 2/(k+3)$ and so Theorem 2 implies that every k-sum-free set $A\subseteq\mathbb{N}$ with $\overline{\mathrm{d}}(A)>1/\rho_2(k)$ is contained in some arithmetic progression B of difference $M<\rho_2(k)$. As we have already observed in the proof of Corollary, M cannot divide $k-1$ and thus $M=\rho_1(k)$ and $B=\{n: n\equiv s \pmod{\rho_1(k)}\}$, where $1\leqslant s\leqslant\rho_1(k)$. From definition of $\rho_1(k)$ it follows that for some prime p and a natural number r we have $\rho_1(k)=p^r$, where $p^{r-1}\mid(k-1)$. Thus, since the fact that B is k-sum-free implies that $\rho_1(k)\nmid s(k-1)$, we have $p\nmid s$, and s and $\rho_1(k)=p^r$ are relative prime. ∎

3. STRONGLY k-SUM-FREE SETS

We start our study of strongly k-sum-free sets of natural numbers with two results analogous to Lemma 1 and Theorem 1.

LEMMA 1*. *Let* $k \geqslant 3$ *and let* $A \subseteq \mathbb{N}$ *be a strongly* k*-sum-free set such that* $x, x+m, ..., x+(i-1)m \in A$ *and for some* $d \in \mathrm{Df}_{\leqslant k}(A)$ *we have* $d \equiv x \pmod m$. *Then*

$$\overline{\mathrm{d}}(A) \leqslant \frac{k+i-1}{2(k-1)\,i+k(k-1)}.$$

Proof. Besides of technical details, the proof of Lemma 1* follows a similar way as the argument which led us to Lemma 1. Thus, let us suppose that A, x, m, and d fulfill the assumption of Lemma 1* and, furthermore, let us assume that $d<x$. Then, for some $d_1 \in \mathrm{Df}_{\leqslant k}(A)$ we have $d_1+m, ..., d_1+im \notin \mathrm{Df}_{\leqslant k}(A)$. Let $d_1 = u - v_1 - \cdots - v_r$ for some r, where $1 \leqslant r \leqslant k-1$ and $u, v_1, ..., v_r \in A$, and let

$$C = \{a \in A : a + jm \notin A \text{ for every } j = 1, 2, ..., i\}$$

Consider the family of sets which consists of

$$C,\ C-x-(x+m), ...,\ C-x-(x+(i-1)\,m)$$

and

$$A-m,\ A-(x+m),\ A-(x+m)-(x+(i-1)\,m), ...,$$
$$A-(x+m)-(x+(i-1)\,m)-\cdots-(x+(i-k+2)\,m),$$

where for $i-j<0$ instead of element $x+(i-j)\,m$ we take x. It follows immediately from the definition of C and the fact that A is strongly k-sum-free that all $i+k$ above sets are disjoint and thus, for large n,

$$k\,\overline{\mathrm{d}}_n(A) = i\,\overline{\mathrm{d}}_n(C) \leqslant 1 + O(1/n). \qquad (**)$$

In order to get the second inequality involving $\overline{\mathrm{d}}(A)$ and $\overline{\mathrm{d}}(C)$ we consider the following four families of sets:

$$\mathscr{A}_1 = \{A, (A \backslash C)+v_1, ..., (A\backslash C)+v_1+\cdots+v_r\}$$
$$\mathscr{A}_2 = \{(A\backslash C)+u+v_1+\cdots+v_r, ..., (A\backslash C)+(k-1-r)\,u+v_1+\cdots+v_r\}$$
$$\mathscr{A}_3 = \{A+u+v_1, ..., A+u+v_1+\cdots+v_{r-2}\}$$
$$\mathscr{A}_4 = \{A+u+v_1+\cdots+v_{r-1}, ..., A+(k-r)\,u+v_1+\cdots+v_{r-1}\},$$

where for $i \leqslant 0$ we omit all shifts of the set A, where element v_i occurs. We shall show that all sets from the family $\{A\} \cup \bigcup_{i=1}^4 \mathscr{A}_i$ are disjoint. From the fact that A is strongly k-sum-free it follows that sets from $\bigcup_{i=1}^4 \mathscr{A}_i$ do not share elements with A and that all families $\mathscr{A}_i$, where $i=1, 2, 3, 4$, consists of pairwise disjoint sets. It is also easy to see that sets from $\mathscr{A}_1$ and $\mathscr{A}_2$ are disjoint, and that sets from $\mathscr{A}_3$ are disjoint from sets from $\mathscr{A}_2 \cup \mathscr{A}_4$.

Thus it is enough to examine pairs of families $\mathscr{A}_1$ and $\mathscr{A}_3$, $\mathscr{A}_1$ and $\mathscr{A}_4$, $\mathscr{A}_2$ and $\mathscr{A}_4$. Since all these cases can be dealt with in a similar way, we shall verify our claim only for the first pair of families, $\mathscr{A}_1$ and $\mathscr{A}_3$.

Let us suppose that for some s and t, where $1 \leqslant s \leqslant r$ and $1 \leqslant t \leqslant r-2$, we have

$$((A \backslash C)+v_1+\cdots+v_s) \cap (A+u+v_1+\cdots+v_t) \neq \emptyset.$$

If $t \geqslant s$ then we immediately arrive at a contradiction with the assumption that A is strongly k-sum-free. Thus, let us assume that $t<s$. Then there exists $a \in A \backslash C$ and $b \in A$ such that

$$a+v_1+\cdots+v_s=b+u+v_1+\cdots+v_t$$

and so

$$d_1=a-b-\sum_{i=1}^{t} v_i-\sum_{i=s+1}^{r} v_i \in \mathrm{Df}_{\leqslant k}(A).$$

Since $a \in A \backslash C$ then there exists j_0 such that $1 \leqslant j_0 \leqslant i$ and $a+j_0 m \in A$. Hence $d_1+j_0 m \in \mathrm{Df}_{\leqslant k}(A)$ which contradicts the choice of d_1.

Thus, all sets from the family $\{A\} \cup \bigcup_{i=1}^4 \mathscr{A}_i$ are disjoint and so, for large n,

$$(2k-1)\,\bar{\mathrm{d}}_n(A)-(k-1)\,\bar{\mathrm{d}}_n(C)=(k+1)\,\bar{\mathrm{d}}_n(A)+(k-1)\,\bar{\mathrm{d}}_n(A \backslash C) \leqslant 1,$$

which together with $(**)$ gives the required upper bound for $\bar{\mathrm{d}}(A)$.

Similarly as in the proof of Lemma 1 one can mimic the above argument for the case $d>x$ with d_1 replaced by $d_2 \in \mathrm{Df}_{\leqslant k}(A)$ chosen in such a way that $d_2-m, ..., d_2-im \notin \mathrm{Df}_{\leqslant k}(A)$ and with

$$C'=\{a \in A : a-jm \notin A \text{ for every } j=1, 2, ..., i\}$$

playing role of C. ∎

THEOREM 1*. *For every $k \geqslant 3$ and $\varepsilon>0$ there exists a natural number $M'=M'(k, \varepsilon)$ such that every strongly k-sum-free set $A \subseteq \mathbb{N}$ with $\bar{\mathrm{d}}(A)>1/(2k-1)+\varepsilon$ is contained in a strongly k-sum-free set which is a union of arithmetic progressions of difference M' each.*

Proof. Let $\overline{\mathrm{d}}(A) > 1/(2k-1) + \varepsilon$, and let $i_0 \in \mathbb{N}$ be chosen in such a way that

$$\overline{\mathrm{d}}(A) > \frac{k+i_0-1}{(2k-1)\, i_0 + k(k-1)}.$$

Furthermore, let $L = L(\varepsilon, i_0)$ be defined as in Lemma 2 and $M' = L!$. Finally, let us put

$$R = \{r \in N\colon \text{there exists } a \in A \text{ such that } r \equiv a \pmod{M'}\}$$

and

$$D = \{r \in R\colon r + t_1 + \cdots + t_s \notin R \text{ for each } t_1, ..., t_s \in R \text{ and } 1 \leqslant s \leqslant k-1\}.$$

Let us suppose that the assertion does not hold, i.e., that R is not strongly k-sum-free. Then, there exists ℓ and $x, x_1, ..., x_\ell \in R$ such that $2 \leqslant \ell \leqslant k$ and $x_1 + \cdots + x_\ell = x$. Now consider $(k-\ell+1)\,\ell + \ell - 1$ sets

$$\begin{aligned}
&D,\ D + x_1,\ D + x_1 + x_2,\ ...,\ D + x_1 + \cdots + x_{\ell-1}\\
&D + x,\ D + x + x_1,\ D + x + x_1 + x_2,\ ...,\ D + x + x_1 + \cdots + x_{\ell-1}\\
&\quad\cdots\\
&D + (k-\ell)\, x,\ D + (k-\ell)\, x + x_1,\ ...,\ D + (k-\ell)\, x + x_1 + \cdots + x_{\ell-1}\\
&D + (k-\ell+1)\, x,\ D + (k-\ell+1)\, x + x_1,\ ...,\\
&\qquad D + (k-\ell+1)\, x + x_1 + \cdots + x_{\ell-2}.
\end{aligned}$$

Note that D is strongly k-sum-free, and thus all above sets are disjoint. Since $(k-\ell+1)\,\ell + \ell - 1 \geqslant 2k-1$ for every $\ell = 2, 3, ..., k$, we have $\overline{\mathrm{d}}(D) \leqslant 1/(2k-1)$, and so $\overline{\mathrm{d}}(A \backslash D) > \varepsilon$.

From Lemma 2 it follows that one can find natural numbers $m \leqslant L$ and x such that $x, x+m, ..., x+(i_0-1)\, m \in A\backslash D$, and so $x, x+m, ..., x+(i_0-1)\, m \in R\backslash D$. From the definitions of sets D and R it follows that $x + t_1 + \cdots + t_r = t$ for some $t, t_1, ..., t_r \in R$ and $0 \leqslant r \leqslant k-1$. Hence there exist $u, v_1, ..., v_r \in A$ such that $u \equiv t \pmod{M'}$ and $v_i \equiv t_i \pmod{M'}$ for every $i = 1, 2, ..., r$.

Thus $d = u - v_1 - \cdots - v_r \in \mathrm{Df}_{\leqslant k}(A)$ and $d \equiv x \pmod{M'}$. Since M' is divisible by m we have $d \equiv x \pmod{m}$ and $x, x+m, ..., x+(i_0-1)\, m \in A$. But Lemma 1* states that then

$$\overline{\mathrm{d}}(A) < \frac{k+i_0-1}{(2k-1)\, i_0 + k(k-1)},$$

contradicting the assumption. ∎

COROLLARY. *Let $k \geqslant 3$. Then for every $\varepsilon > 0$ there exists $M' = M'(k, \varepsilon)$ such that for every strongly k-sum-free set $A \subseteq \mathbb{N}$ which contains multiplicity of n for every $n = 1, 2, ..., M'$, we have $\overline{\mathrm{d}}(A) \leqslant 1/(2k-1) + \varepsilon$. In particular, if a strongly k-sum-free set $A \subseteq \mathbb{N}$ contains multiplicities of every natural number, then $\overline{\mathrm{d}}(A) \leqslant 1/(2k-1)$.*

Note that also in this case a simple example similar to that mentioned for k-sum-free sets shows that the above result is sharp. Indeed, for an irrational number α the set

$$E_{\alpha, \leqslant k} = \left\{ n \in N : \alpha n - [\alpha n] \in \left(\frac{1}{2k-1}, \frac{2}{2k-1} \right) \right\}.$$

is strongly k-sum-free, $\overline{\mathrm{d}}(E_{\alpha, \leqslant k}) = 1/(2k-1)$ and $E_{\alpha, \leqslant k}$ contains a multiplicity of every natural number.

Unfortunately, at this moment we are not able to prove a result analogous to Theorem 2 because we do not know any good estimate of the size of the set $\bigcup_{i=1}^{k} (iA)$ for a k-sum-free subset A of an abelian group. Thus, we conclude the note with a somewhat weaker statement which says that every strongly k-sum-free set A with upper density larger than $1/(k+1)$ is a subset of an arithmetic progression of difference k.

Let us start with a series of simple consequences of Lemma 1*. Here and below we assume that A is a strongly k-sum-free set for some $k \geqslant 3$ and set

$$m = m(A) = \min\{x - y > 0 : x, y \in A\}.$$

CLAIM 2. *If $m \neq k$, then $\overline{\mathrm{d}}(A) \leqslant 1/(k+1)$.*

Proof. If $m > k$ then clearly $\overline{\mathrm{d}}(A) \leqslant 1/m \leqslant 1(k+1)$. Thus let us assume that $m < k$ and for some $x \in A$, we have $x + m \in A$. Put $u = x + m$, $v_1 = \cdots = v_m = x$. Then $d = u - v_1 - \cdots - v_m \in \mathrm{Df}_{\leqslant k}(A)$ and $d = x + m - mx$, so $d \equiv x \pmod{m}$. Now it is enough to apply Lemma 1* for $i = 2$. ∎

CLAIM 3. *If $x_1, ..., x_r \in A$ are such that $1 \leqslant r \leqslant k-1$ and $x_1 + \cdots + x_r \equiv 0 \pmod{k}$, then $\overline{\mathrm{d}}(A) \leqslant 1/(k+1)$.*

Proof. Because of Claim 2 it is enough to consider the case when $m = k$, i.e. when $x, x + k \in A$ for some $x \in A$. Let us set $u = x$, $v_i = x_i$ for every $i = 1, 2, ..., r$. Then $d = u - v_1 - \cdots - v_r \in \mathrm{Df}_{\leqslant k}(A)$ and $d \equiv x \pmod{k}$, so from Lemma 1* we get $\overline{\mathrm{d}}(A) \leqslant (k+1)/(k^2 + 3k - 2) \leqslant 1/(k+1)$. ∎

CLAIM 4. *If for some $a \in A$ we have $(k, a) > 1$, then $\overline{\mathrm{d}}(A) \leqslant 1/(k+1)$.*

Proof. Since $(k, a) > 1$ there exists a natural number r such that $r < k$ and $ra \equiv 0 \pmod{k}$. Set $x_i = a$ for every $i = 1, 2, ..., r$. Then $x_1 + \cdots + x_r \equiv 0 \pmod{k}$, and Claim 3 implies that $\overline{\mathrm{d}}(A) \leqslant 1/(k+1)$. ∎

Claim 5. *If there exist* $a, b \in A$ *such that* $a \not\equiv b \pmod{k}$, *then* $\overline{\mathrm{d}}(A) \leqslant 1/(k+1)$.

Proof. From Claim 4 it follows that we may assume that $(a, k) = 1$ and $(b, k) = 1$. Moreover, due to Claim 2, it is enough to examine the case when $m = k$, i.e., when for some $x \in A$ we have $x + k \in A$. Without loss of generality we may assume that $x \not\equiv a \pmod{k}$. Furthermore, because of Claim 3 it is enough to study the case when $x \not\equiv 0 \pmod{k}$. Since $(a, k) = 1$, all numbers $a, 2a, ..., ka$ give distinct reminders from division by k and for some $r < k$ we have $ra \equiv a - x \pmod{k}$. Now if we set $u = v_1 = x$ and $v_i = a$ for every $i \in \{2, ..., r\}$, we get $d = u - v_1 - \cdots v_r \in \mathrm{Df}_{\leqslant k}(A)$ and $d \equiv x \pmod{k}$, so the assertion follows from Lemma 1* with $i = 2$. ∎

Thus we arrived at the following result.

Theorem 3. *If* $k \geqslant 3$ *and* $A \subseteq \mathbb{N}$ *is a strongly k-sum-free set with* $\overline{\mathrm{d}}(A) > 1/(k+1)$ *then there exists number s, such that* $(s, k) = 1$ *and* $A \subseteq \{n : n \equiv s \pmod{k}\}$.

Proof. From Claim 5 it follows that $A \subseteq \{n : n \equiv s \pmod{k}\}$, while Claim 4 implies that $(s, k) = 1$. ∎

Clearly, since the set $\{n : n \equiv 1 \pmod{(k+1)}\}$ is strongly k-sum-free, the constant $1/(k+1)$ in the above result cannot be replaced by a smaller one.

REFERENCES

1. N. J. Calkin and P. Erdős, On a class of aperiodic sum-free sets, *Math. Proc. Camb. Phil. Soc.*, to appear.
2. J.-H. Deshouillers, G. Freiman, V. Sós, and M. Tamkin, On the structure of sum-free sets 2, to appear.
3. P. Erdős, Extremal problems in number theory, *Proc. Symp. Pure Math.* **8** (1965), 181–189.
4. T. Łuczak, A note on the density of sum-free sets, *J. Combin. Theory Ser. A* **70** (1995), 334–336.
5. J. H. B. Kemperman, On small sumsets in abelian group, *Acta Math.* **103** (1960), 63–88.
6. M. Kneser, Ein Satz über abelschen Gruppen mit Anwendungen auf die Geometrie der Zahlen, *Math. Z.* **61** (1955), 429–434.
7. T. Łuczak and T. Schoen, On large sum-free subsets of abelian groups, submitted.
8. E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, *Acta Arith.* **27** (1975), 299–345.